Pioneers in Proteomics

Joshua LaBaer, M.D., Ph.D.
Director, Harvard Institute of Proteomics
Harvard Medical School

In part three of our Pioneers in Proteomics series, Dr. Joshua LaBaer discusses the factors critical to advancing the field of proteomics and the challenges of applying proteomics to cancer diagnosis and treatment.

Dr. LaBaer is Director of the Harvard Institute of Proteomics at Harvard Medical School. He is currently an Attending Physician at the Dana-Farber Cancer Institute and holds an academic appointment through the Department of Biological Chemistry and Molecular Pharmacology at Harvard Medical School. Together with Dr. Ed Harlow, Dr. LaBaer founded the Harvard Institute of Proteomics in the spring of 1999.

1. On the hurdles to overcome

So there are a number of hurdles that are blocking getting proteomics more into clinical research. And they fall into several different categories, and I think all of them, at some level, are going to be helped by the Clinical Proteomics Technologies Initiative.

So the first is we need to get more clinical samples. Because in order to identify markers for disease, we need to look at a lot of clinical samples. That's the only way to be able to generate the statistics to say that this sample -- that this marker is clearly predictive of disease, and not just random variation among different individuals. And so for that we need well-defined clinical samples. They need to be well annotated clinical samples. And they need to be made available to lots of people

The second key area that's needed is technology development. Right now, technology is still fairly crude, in terms of our ability to look at the deep dynamic range of proteins in serum, that is to say, you know, the most rare proteins, compared to the most abundant proteins, and being able to find the most rare proteins. And so we need better technologies for that. We also -- we need better and more creative technologies for looking at protein function. Being able to look at what roles proteins might play in cancer, either in terms of pathways that they participate in, or studying simple protein catalytic activity on chips, or looking at protein networks, or protein-protein interactions. We need more reagents that we can use to study protein function. This includes antibodies to proteins. These antibodies will not only be useful in actual biological research, but they will also be something that someone could imagine putting on a protein chip to look at protein abundance in various specimens.

And then the last thing is by creating a centralized place people can go to understand

what are the best standard operating procedures that we should be using to do research? So if we all have a common -- common SOPs that we use to collect clinical samples, common SOPs that we use to separate proteins prior to loading them on an instrument, that will allow us all to speak the same language, and cross-compare information from one laboratory to another.

2. On the need for creative technology development

I think the other thing that we need is just more creativity. We need to be thinking about other technological approaches, and other ways of looking at proteins, other than just simply measuring protein abundance. This is where I think protein micro arrays will play an important role, because it allows us to look at protein function. By identifying which proteins play roles that are likely to link them to cancer processes, we have a better chance of identifying proteins that might also themselves be markers, or targets for disease.

And this is where I think programs that help encourage scientists to do technological development is key. Right now it's typically very difficult for scientists to develop new technologies, because most granting mechanisms focus on hypothesis-driven research. I think what we need to -- we need to, at least on some level, fund scientists to do technology development, with the understanding that there has to be a clear goal as to how that technology's going to impact, you know, translational research.

3. On finding the "needle in the haystack"

One of the other limitations that has kept proteomics from making a big impact so far on the clinic is that the technologies that are available right now for proteomics are not quite ready for some of the technical challenges that they're faced for doing clinical samples -- for evaluating clinical samples. It's been harped on a lot by people who do proteomics, and everyone's aware of it. But it still remains a huge problem that the proteins that we're looking for that are likely to to be good diagnostic markers are in serum, or in a solution that has an enormous dynamic range, in terms of the amount of proteins that are there. This makes it very hard to identify those proteins at the level that we need to look at, with all the much more abundant proteins that are there. It's very much the needle in the haystack problem**.** And newer technologies, and improved technologies are going to be necessary to be able to find those proteins.

The issue of dynamic range for identifying markers in serum is that from the most abundant proteins in serum to the least abundant protein system in serum, there are ten orders of magnitude. That's ten, times ten, times ten, times you know, all the way up. So this means that there are some proteins in the serum that are so rare that they're overwhelmed by the other proteins that are present in the serum. There's lots and lots of other things present. But the difference between the most abundant and the least abundant is very, very large. And being able to find, unfortunately, the proteins that we're interested in that predict disease tend to be the more rare proteins. So we need to find a way to not look at all the abundant stuff, and get down to the much more rare stuff.

What makes it even more challenging is that because those proteins are hard to find, even when you find them, you need to prove that they're predictive. And this gets back to the problem of having to look at many different patients, and look at multiple clinical samples to prove that there is at least specificity there in that protein, if not some sensitivity, in terms of identifying a marker for a disease.

4. On the importance of standards

When we use clinical samples, working with SOPs becomes very important. It's important that everybody collects samples in a similar way, or we can't compare results from one laboratory to another. Within a laboratory the processing that then goes on needs to follow SOP so that the results from that laboratory conclusions can be made from. But we also have to be careful not to box the entire community in too much, with respect to insisting that every lab do its experiments exactly like every other lab, or else what we'll find is we'll shut down, to some extent, the discovery that may happen in the different laboratories. To some extent, having a little bit of variation helps in the discovery.

Obviously, to a large extent, NIH can play an important role there. Certainly in the cancer field NCI has already been playing an important role there. In part, by making sure that everybody comes together on a regular basis, and discusses what they're doing, and how they're doing it.

5. On the need for multidisciplinary collaboration

I think one of the key things about proteomics that's becoming more and more apparent is that the more shared information that we have out there, the better we are at identifying proteins**.**

And one of the advantages of the clinical proteomics initiative is to essentially enforce

and allow that kind of sharing of information. It means that different laboratories will be able to look at the results of other laboratories. To be able to interpret them in their own ways, using their own software, and be able to better understand what they're looking at.

Some of the key outcomes of the Human Genome Project, probably perhaps the most valuable outcomes, was helping us pinpoint what genes, and ultimately what proteins are there. And also changing a little bit the culture of biological sciences, which for a long time had been small, focused laboratories, working in very directed ways. And the Genome Project really opened up everyone's mind to the idea of forming large teams of scientists to study proteins -- or to study biology in a multidisciplinary way. So bringing together informaticists, and engineers, and mechanical engineers, biologists, molecular biologists. And bring them together into a team environment to start studying biological questions.

And I think the same is going to be true for studying clinical proteomics. That it's clear that we're going to need people who are instrument specialists, and who are very knowledgeable about protein separation. People who understand how to write software to identify proteins, or people who know how to work with various technologies that they've developed in their laboratories. But we also need to involve clinicians who understand the diseases, and understand the variations that are natural in the diseases.

6. On partnering with clinical researchers

It's clear that proteomics is being applied to clinical questions. I think one of the other limitations of proteomics right now, and clinical applications, is that by and large a lot of the folks who do proteomics are not necessarily well-versed in clinical research themselves. And the ones who do the best job are the ones who pair up with clinical researchers. And I think this gets back to the whole team approach that's needed in proteomics. What's needed is having people who think about epidemiology, who think about biostatistics. Who understand how to do, you know, sensitivity specificity, ROC curve studies, understand issues around oversampling, and overfitting that apply when you study thousands of samples at the same time. That is stuff's not something a person who's used to doing mass spectrometry or who builds protein microarrays thinks about all the time. We need to get together with our clinical partners, and work with them, as well as our biostatisticians, and make sure that we're all doing experiments that we need.

7. On the use of biomarkers in cancer

In cancer, some markers may be present in subsets of patients with a particular cancer. You could imagine with breast cancer, for example, that a subset of women with breast cancer may produce a particular biomarker, but not all women. And therefore, the sensitivity of that marker may not be high. Now what we need to -- what we'll definitely need are markers that have high specificity**.** We'll need to know that if that marker's present that -- or when I say present I mean present at whatever level it is that we define as the differential between calling it a positive test and a negative test. That we know that whenever that test is positive that it's strongly predictive of cancer, as opposed to sort of a weak marker. And then, by putting together a panel of multiple proteins, we may be able to catch those women who may not have that first marker, but may produce a second marker, or a third marker. And then the hope is that by bringing together multiple -- multiple proteins or peptides, that we'll improve both the sensitivity and the specificity of the test.

It's becoming more and more apparent that there are no single biomarkers that have enough sensitivity, and enough specificity to diagnose cancer, for example. And I think that's probably going to be true for other diseases, ranging from diabetes, to autoimmune diseases, and what not. And so more likely than not we will identify panels of proteins. But we'll still need to know the specifics of each member of that panel. So I distinguish a panel of proteins from a pattern. A pattern might be just some descriptive term that could be used as a marker for an illness. But ultimately I think to get it to the clinic we're going to need to identify all the components of that pattern, understand each of the proteins that are being measured, and really learn the values. And that will take the pattern to a panel, and we'll understand the sensitivity and specificity of each member of that panel, for example.

8. On the benefits to cancer patients

How will proteomics be used ten years from now? The first is I think the impact of current proteomics will be to identify better and more markers that can be used to follow various diseases. The second is that we all now recognize that diseases -- diseases that we had thought about as being very simple, or at least very -- falling into one single category are themselves heterogeneous. So, for example, you know, adenocarcinoma of the breast may actually have multiple types of adenocarcinoma, or adenocarcinomas that have different prognoses associated with them. And proteomics may play an important role in allowing us to look at a patient who walks in the clinic and saying, "This one needs chemotherapy, and this woman does not need chemotherapy, because she already

has a disease based on this proteomic pattern," for example, "that is unlikely to metastasize. And therefore, simple surgical removal and radiation therapy may be satisfactory for that illness." Or same thing may be true for prostate cancer. We already know that a huge fraction of men in their seventies have some degree of prostate cancer, but only a relatively smaller fraction of that actually end up going on to metastatic disease, and dying from it. And so wouldn't it be useful if we could take out a tumor specimen, evaluate the proteins in that specimen, and then be able to say, "This is a type of prostate cancer unlikely to metastasize to bone, or to -- this is a type of disease that might be simply treatable by, you know, surgical removal, or even may not require surgical removal."

9. On the promise of proteomics

I think that right now proteomics is very much still in the discovery phase. I think we still don't know what the best markers are, and we're all working very hard to identify them. And I say markers, but I think proteomics is much bigger than just markers. To a large extent, ultimately proteomics will also help us identify targets for therapeutics. They'll also identify pathways that play important roles. Not only in treating disease, but also in treating resistance to other drug therapies. We may find, as we've known for a long time in cancer treatment, that combined therapies are key to getting success. And so we're going to need multiple pathways to treat if we're going to be successful.